

## **Method Of Providing Quality Of Service (QOS) To Voice Applications In Routed IP Networks.**

### **FIELD OF THE INVENTION**

The present invention relates generally to the field of Internet Protocol (IP) networks, and more specifically to the field of providing Quality Of Service (QOS) within such networks.

### **BACKGROUND OF THE INVENTION**

5           Routers in existing IP networks do not currently accommodate the reserving of bandwidth for traffic management purposes. Accordingly, a router has no up-to-date notion of what is occurring with its associated endpoints in terms of accepting new traffic. For instance, once certain VoIP traffic has been admitted, there is no easy way for the router to determine that the specific traffic is no longer continuing. Thus, packets are typically dropped if the router cannot handle the associated traffic. One possible solution to address this problem is to associate timers to the specific traffic for each connection on each of the links and monitor each call for activity. Such a methodology is extremely costly, however, in that multiple timers must be programmed into each router and tracked. Accordingly, there is a need for a simplified traffic management methodology for routers in IP networks.

### **SUMMARY OF THE INVENTION**

A simplified methodology for accomplishing traffic management in a packet based network is achieved by allowing allocation of bandwidth based on a count of the number of endpoint connections associated with a specific service for a corresponding network device. A network device (e.g., IP router) can be configured to count the number of TCP/IP (Transmission Control Protocol/Internet Protocol) and/or UDP/IP (User Datagram Protocol/Internet Protocol) connections and bandwidth usage per interface. The counting can be done by identifying the IP Addresses, Type of Service (TOS), and TCP/UDP and/or UDP/IP socket number range in the IP Header of a packet and then querying a specific communication type. When reaching the maximum allowed connections or bandwidth for a specific service, the network device (e.g., IP router) stops

forwarding any new calls by means of dropping packets of new calls and informing the given endpoints to disconnect the new calls.

In one exemplary embodiment of the invention, a procedure for managing traffic flowing through individual routers of a packet network includes the steps of reserving a given amount of bandwidth on interfaces of the individual routers for specific types of communications traffic, periodically querying endpoint connections based on data from a corresponding router connection table, receiving responses from the periodic querying to determine a current connection status and bandwidth allocation of said endpoints and calculating current bandwidth allocation for a specific type communications service on an interface handled by the router. The router admits additional communications traffic for a specific type of communications service if bandwidth is available.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

A more complete understanding of the present invention may be obtained from consideration of the following detailed description of the invention in conjunction with the drawing, with like elements referenced with like references, in which:

FIG. 1 is an exemplary embodiment of a router in an IP network which makes use of the present invention;

FIG. 2 is an exemplary embodiment of an endpoint connection table for an associated network;

FIG. 3 is an exemplary embodiment of an IP network used in connection with the present invention;

FIG. 4 is an exemplary block diagram of an external monitoring device in accordance with the present invention; and

FIG. 5 shows an additional embodiment of an IP network having calls routed through a VoIP gateway.

### **DETAILED DESCRIPTION**

The present invention enables allocation and tracking of bandwidth based on a count of the number of specific types of connections, e.g., voice over IP calls, currently in existence. In accordance with the present invention, a network device (e.g., IP router) is configured to count the number of TCP/IP (Transmission Control Protocol/Internet

Protocol) and/or UDP/IP (User Datagram Protocol/Internet Protocol) connections per interface. The counting is done by identifying the IP Addresses, Type of Service (TOS), and TCP/UDP and/or UDP/IP socket number range in the IP Header of a packet. A table in the router is populated based on the information provided by the count and the allocated bandwidth is totaled. When reaching the maximum allowed bandwidth that will be accommodated, the network device (e.g., IP router) stops forwarding any new calls. In an exemplary embodiment of the invention, this stoppage is accomplished by way of dropping packets of new calls and informing the given endpoints to disconnect the new calls.

Referring to Fig. 1, one embodiment of an IP router 10 is shown in accordance with the present invention. As can be seen, the router includes a number of interfaces, interfaces A, B, C. Interface A couples to VoIP gateway 12 which in turn couples to the public switched telephone network (PSTN). H.323 gateways 12 provide services to H.323 clients so that they can communicate with non H.323 entities. The most common type of H.323 gateways allow communications between H.323 terminals and telephones on the circuit switched network. The gateway must provide translations between different transmission formats, communications procedures and audio codecs. Interface B couples to one or more endpoint IP phone/PC phone terminal(s) 16, interface B is shown to have a H.323 real time traffic flow to a first endpoint terminal. The endpoint terminal 16 provides real time communications and support voice communications and can optionally support video and data communications. The most common H.323 terminals are applications, such as Microsoft's NetMeeting, running on a PC.

In connection with the present invention, the IP router 10 includes software stored in memory (not shown) which enables the router to periodically count or ascertain the number of active connections per interface. The counting is based on TOS, IP Addresses, and TCP/UDP and UDP/IP socket range. When a router receives a new packet carrying a new IP address a connection is able to be identified and the router assigns a priority to the flow based on the TOS (ascertained from the packet header) and/or TCP/UDP socket. For example, if the flow is a non H.323 data query, the router will assign a low priority to the flow in its connection table. However, if the flow is an H.323 real time traffic flow (e.g.,

VoIP), the flow will be treated differently. A real time flow, for example, VoIP or video will be assigned to the highest priority in the connection table.

The present invention enables maintaining an up to date status of bandwidth allocation through periodic updates of the connection table in the corresponding router.

5 In order to update the connection table, the IP router periodically (e.g., at 10 second intervals) sends an H.323 IRQ status query message to the H.323 endpoints in its connection table. The router will then set the Call\_Ref value of the IRQ message to 0. The H.323 endpoint will respond with an IRR message to inform the router about the status of all of its calls (e.g., active or inactive) and the bandwidth utilized by the call. A  
10 table is created and maintained in the router for each interface so that a total of the currently allotted bandwidth may be readily updated and accessed. Referring to Fig. 2, an exemplary embodiment of a router call connection table (Table II) is shown for Interface 1 of router R1 in the network 200. The table includes various categories for each connection of a respective interface. As shown, the table includes a category for Source  
15 IP Address, Call Reference, Utilized Bandwidth and Connection Status. The table also illustrates that each call connection of an associated gateway 202 may be tracked separately. In accordance with the present invention, the router may also operate so that queries to the endpoints are made only to the real-time or otherwise highest priority connections.

20 For the response part of the query, if the router has no calls active for the specific query, the router will set the Call\_ref value of the IRR message to 0 indicating that it has no active calls. In this case, the router will update its table by deleting the entry belonging to the call and decrementing the number of H.323 connections in the connection table. When the number of H.323 calls reaches the maximum, the router will drop packets of  
25 any new connections.

When the number of calls or the bandwidth allotment for a router interface reaches its maximum, the router now can do one of two things to inform the endpoint to disconnect the call if the maximum number of calls is reached/exceeded. A first approach is to do nothing and continue dropping packets of new calls. In this approach, the  
30 endpoints will realize that they are experiencing 100% packet loss (e.g., by the RTPCP

messages interchanged between the endpoints). In this case the higher layers in the terminals will terminate the call because it experiences too much packet loss.

In the second approach, the Router first obtains the Call\_Ref value of the new call. It will then construct a new message by using the non-Standard H.245 message format and transmit it to the endpoint to inform it that to terminate the call. This new message will have the following information:

Name Network Congestion Disconnect (NCD)

Call\_Ref       xx

Referring to Fig. 3, there is shown another embodiment of an IP network 20 which utilizes the present invention. As shown, the network 20 includes four routers, IP router 1, 2, 3, 4. Each of the IP routers 1, 2, 3, 4 are part of the overall IP network 20 and may include connections to one or more H.323 endpoints. As shown, router 1 and router 3 couple to a first H.323 endpoint 22 (e.g., an IP phone or PC phone) and router 2 and router 3 each couple to a second H.323 endpoint 24. The IP network couples to the PSTN network 26 and voice switches 28 through a VoIP gateway 30 that couples to router 4. In addition, although the endpoints shown in Figs. 1 and 2 are described as IP or PC phones, it would be understood that the endpoints may also be endpoints with video and/or multimedia capabilities.

Fig. 3 illustrates that an optional external device 32 can also be responsible for querying the associated endpoints of a router, counting the number of connections and associated bandwidth utilization, and blocking new connections when capacity is reached, as in the case of the router. The basic structure of the external device 32 is shown in Fig.4. As shown, the device includes a digital processor 34 and associated memory 36 for storing the count, query and update program for execution by the processor. A communications interface 38 for communicating with the endpoints and corresponding router(s) is also included.

Referring to Fig. 5, an alternate network configuration 50 having routers R1, R2, R3 for use in connection with the present invention is shown. Fig. 5 illustrates that multiple IP telephone (or terminal) devices 52 can be coupled to an IP network through a single VoIP gateway. With such a configuration, a single IRQ query to the gateway 56 will provide a snapshot (of all the active calls together with information about the

bandwidth of each call) of the VoIP or other specific bandwidth which is allocated through the gateway. Such a configuration is extremely efficient, for providing multiple connection information back to the router. Implementation of the present invention is straightforward in that, the IRQ and IRR message formats, along with a methodology for counting need only be programmed into the routers or corresponding external device.

The present invention advantageously makes use of H.323 query and response messages which are currently available in the H.323 protocol, i.e., IRQ and IRR messages. In the past, the IRQ command has been sent from a network management device or gatekeeper to a terminal requesting status information in the form of an IRR.

Thus, the commands have been used in the context of network management tasks, but not in connection with routers for traffic management purposes.

This invention is advantageous in that it provides a mechanism to guarantee QoS in VoIP networks through what is in effect dynamic reservation of bandwidth per call. This, in turn, allows for traffic engineering, e.g., allowing over-subscription of Voice over IP calls in IP networks, and therefore reduces the amount of data links required. Prior to this invention, in order to implement a VoIP network, all data links had to be under-subscribed.

In order to ensure accurate bandwidth allocation statistics, the querying process of the present invention can be periodically reset at predetermined intervals (e.g., every five minutes) or in response to specified events occurring within the network (e.g., router failures or link failures).

The foregoing description merely illustrates the principles of the invention. It will thus be appreciated that those skilled in the art will be able to devise various arrangements, which, although not explicitly described or shown herein, embody the principles of the invention, and are included within its spirit and scope. Furthermore, all examples and conditional language recited are principally intended expressly to be only for instructive purposes to aid the reader in understanding the principles of the invention and the concepts contributed by the inventor to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions.

Moreover, all statements herein reciting principles, aspects, and embodiments of the invention, as well as specific examples thereof, are intended to encompass both structural

and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents as well as equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure.

In the claims hereof any element expressed as a means for performing a specified  
5 function is intended to encompass any way of performing that function including, for example, a) a combination of circuit elements which performs that function or b) software in any form, including, therefore, firmware, microcode or the like, combined with appropriate circuitry for executing that software to perform the function. The invention as defined by such claims resides in the fact that the functionalities provided by  
10 the various recited means are combined and brought together in the manner which the claims call for. Applicant thus regards any means which can provide those functionalities as equivalent as those shown herein. Many other modifications and applications of the principles of the invention will be apparent to those skilled in the art and are contemplated by the teachings herein. Accordingly, the scope of the invention is limited  
15 only by the claims appended hereto.